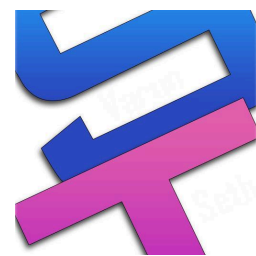




TAKNEEK PS - On Spot

(50 Points)



Recognition of Adversarial Covert Channel Outliers in Operational Networks

Introduction :

You are a team of elite cyber threat analysts at a major **financial** institution. For the past 48 hours, our network sensors have collected vast amounts of network flow and DNS query data. We suspect a sophisticated adversary has compromised several internal machines and is operating a stealthy botnet. Your mission is to sift through this **mountain** of data, and pinpoint the compromised hosts and the external C2 servers they are communicating with.

Your goal is to identify and report two sets of IP addresses from the provided 48-hour dataset:

1. The list of internal hosts that are compromised.
2. The list of external C2 server IPs that the bots are communicating with.

Dataset:

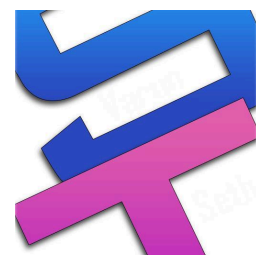
You will be given three CSV files:

1. [network_flows.csv](#): Contains records of network connections (timestamp, source_ip, dest_ip, dest_port, protocol, bytes_sent, bytes_received). This includes both internal and external traffic.
2. [dns_logs.csv](#): Contains records of DNS queries made by internal hosts (timestamp, client_ip, query_name, response_ips).
3. [host_info.csv](#): Provides some context on a subset of internal IPs (ip_address, role).



TAKNEEK PS - On Spot

(50 Points)



Deliverables and Submission Format:

1. Compressed folder with code files. These should include:
 - a. A README with detailed explanation of your approach/algorithm.
Explain design choices (why a given mapping algorithm was used over the other, why a particular form of feature engineering was implemented?) Each step of the algorithm should be thoroughly explained.
 - b. The README should also contain the instructions to run the inference pipeline. Use relative path for input files in the inference code.
 - c. Code files for training and inference in a well structured format. If there are multiple subfolders then provide a README for each subfolder for clarity.
2. submission.csv file with two columns: ip_address and label ('bot', or 'c2_server').

Evaluation Metric: Macro averaged F1 Score

Rules and Team Composition :

1. At least 2 Y24 participants
2. At most 1 participant from (Y22 + Y23) batch

For Any Queries, The Pool Captains and PS Leads are encouraged to ask in the Discord channel.